Supporting Materials and Methods

Crosslinking of Cells, Fragmentation, and Immunoprecipitation of Chromatin

Formaldehyde Crosslinking. Cells (10^9) suspended in growth medium are transferred as 40-ml aliquots into 50-ml tubes and placed on ice for 10 min. Then 1/10 volume of crosslinking solution (11% formaldehyde/0.1M NaCl/1mM Na-EDTA/0.5 mM Na-EGTA/50mM Hepes, pH 8.0) is added directly to each tube. After 10-min incubation on ice, 1/20 vol of a 2.5-M glycine solution is added to each tube to stop the crosslinking reaction. The cells are harvested by centrifugation at 2,000 × *g* for 10 min at 4ºC. The cell pellets are resuspended in cold PBS (137 mM NaCl/2.7 mM KCl/10 mM Na_2HPO_4/2 mM KH_2PO_4) and washed twice. The final cell pellet is snap frozen in liquid nitrogen.

Extraction of Chromatin. The frozen cell pellet is resuspended in 30 ml of lysis buffer 1 [0.05 M Hepes-KOH, pH 7.5/0.14 M NaCl/1 μM EDTA/10% glycerol/0.5% NP-40/0.25%, Triton X-100, with protease inhibitor cocktail from Roche Applied Science (catalogue no. 1836170)] with a pipette and mixed for 10 min at 4°C on a rocking platform. After centrifugation at 2,000 × *g* for 10 min at 4°C, the cell pellet is resuspended in 24 ml of lysis buffer 2 (0.2 M NaCl/1 μM EDTA/0.5μM EGTA/10 μM Tris, pH 8/protease inhibitor cocktail) with a pipette and mixed gently at room temperature for 10 min on a rocking platform. After centrifugation again at 2,000 × *g* for 10 min at 4°C, the pellet is resuspended in 10 ml of lysis buffer 3 (1 μM EDTA/0.5 μM EGTA/10 μM Tris·HCl, pH 8/protease inhibitor cocktail).

Fragmentation of Chromatin. The mixture is divided into 5-ml aliquots and placed into 15-ml tubes. These tubes are then placed into 50-ml tubes containing ice. An ultrasonic sonicator (Branson Sonifier 450, with power setting at 5) is used to break down the cell and nucleus membranes and fragment the chromatin. The sonicator probe tip is first immersed in the mixture followed by 25 s of continuous sonication. Subsequently, the tube is placed on ice for at least 1 min to avoid accumulation of heat. The sonication is repeated until the chromatin fragments are of desired length. The chromatin fragment size can be examined by agarose gel electrophoresis analysis of 10 μl of cell extracts, digested with protease K for 1 h. The number of sonication cycles varies with different cell types and crosslinking conditions, and pilot tests are recommended. Usually, 10 cycles of sonication are necessary to achieve the desired fragmentation size. Finally, the chromatin solution is adjusted to 0.5% Sarkosyl (sodium lauryl sarcosine) and gently mixed for 10 min at room temperature on a rocking platform. The chromatin solution is then transferred to a centrifuge tube and spun for 10 min at 10,000 × *g* to remove cell debris. The supernatant is collected for chromatin immunoprecipitation. At this step, the DNA concentration of the solution should be around 1-2 mg/ml. The solution can be stored at –80°C as 1-ml aliquots.

Immunoprecipitation of Chromatin. Magnetic beads (Dynal, Oslo) precoupled to the primary antibodies are used to immunoprecipitate the DNA associated with the protein of interest. To prepare the magnetic beads, 100 μl of sheep anti-rabbit IgG-conjugated

dynabeads (Dynal; catalogue no. 112.04) or goat anti-mouse IgG-conjugated beads (Dynal; catalogue no. 110.05) are first washed three times with cold PBS containing 5 mg/ml BSA (Sigma; catalogue no. A-7906) and then resuspended in 5 ml of cold PBS. Rabbit polyclonal (or mouse monoclonal) antibody (10 μg) is added to the mixture and incubated overnight on a rotating platform at 4°C. After collecting the magnetic beads by centrifugation and washing three times with cold PBS containing 5 mg/ml BSA, the beads are resuspended in 100 μl of cold PBS with 5 mg/ml BSA and are ready for immunoprecipitation.

In an Eppendorf tube, 2 mg of soluble chromatin solution is first adjusted to 0.1% Triton X-100, 0.1% sodium deoxycholate, and 1 mM PMSF, then mixed with 100 μl of magnetic beads precoupled with the antibody. The mixture is incubated at 4°C overnight in a rotating platform. The magnetic beads are then collected using a magnet MPC-E (Dynal), and the supernatant is removed by aspiration. To remove materials nonspecifically bound to the beads, 1 ml of RIPA buffer (50 mM Hepes, pH 7.6/1 mM EDTA/0.7% DOC/1% NP-40/0.5 M LiCl/protease inhibitor cocktail) is added to the tube, and the beads are gently resuspended on a rotating platform in a cold room. The magnetic beads are again collected with MPC-E. After washing with RIPA buffer for a total of five times, and once with 1 ml of TE buffer, the beads are collected by centrifugation at 2,000 × *g* for 3 min and resuspended in 50 μl of elution buffer (50 mM Tris, pH 8/10 mM EDTA/1% SDS). To elute precipitated chromatin from the beads, the tubes are incubated at 65°C for 10 min with constant agitation, then centrifuged for 30 s at 2000 × *g*. Supernatant (40 μl) is taken out and mixed with 120 μl of TE buffer with 1% SDS. This solution is incubated at 65°C overnight to reverse the crosslinks. As a control, 100 μg of chromatin is mixed with 120 μl of TE with 1% SDS in a separate tube and is incubated at 65°C overnight.

Purification of DNA. After reversal of crosslinks, proteins in the DNA sample are removed by incubation with 120 μl of proteinase K solution (2% glycogen/5% 20 mg/ml proteinase K stock solution/TE buffer) for 2 h at 37°C. The sample is then extracted twice with phenol (Sigma; catalogue no. P-4557), once with 24:1 chloroform/isoamyl alcohol (Sigma; catalogue no. C-0549). The sample is adjusted to 200 mM NaCl. After ethanol precipitation, the DNA is dissolved in 30 μl of TE buffer containing 10 μg of DNase-free RNase A (Sigma; catalogue no. 6513) and incubated for 2 h at 37°C. The DNA at this step can be further purified with Qiagen PCR kit (Qiagen, Valencia, CA; catalogue no. 28106).

Blunting, Ligation of Linker to DNA, and Amplification by PCR

The immunoprecipitated DNA is usually in so small a quantity (1-10 ng of total) that it is necessary to amplify them significantly for the subsequent labeling step and DNA microarray analysis. To achieve this, a ligation-mediated PCR (LM-PCR) procedure is used. Because the DNA at this step contains uneven ends as a result of the physical shearing process, it is first treated with T4 DNA polymerase to form blunt ends. Then a linker is ligated to the DNA fragments. The added linker allows the DNA to be amplified by PCR using a universal oligonucleotide primer.

Blunting. In an Eppendorf tube, the 40 µl of immunoprecipitated DNA (or 20 ng of control input DNA) is mixed with 11 µl of (10×) T4 DNA polymerase buffer (New England Biolabs; catalogue no. 007-203), 0.5 µl of BSA (10 mg/ml) (New England Biolabs; catalogue no. 007-BSA), 0.5 µl of dNTP mix (20 mM each), 0.2 µl of T4 DNA pol (3 units/µl) (New England Biolabs; catalogue no. 203L), and dH_2O to a total volume of 112 µl. After 20 min incubation at 12°C, 1/10 vol of 3 M sodium acetate (pH 5.2) and 1 µg of glycogen (Roche Applied Sciences; catalogue no. 0901393) is added to the tube, and the DNA sample is extracted with phenol:chloroform:isoamyl alcohol (25:24:1) (Sigma; catalogue no. P-3803) once. After ethanol precipitation, the final DNA is dissolved in 25 µl of dH_2O.

Ligation. The blunt-ended DNA is mixed with 8 µl of dH_20, 10 µl of 5× ligase buffer (Invitrogen; catalogue no. 46300-018), 6.7 µl of annealed linkers (oligo-1: GCGGTGACCCGGGAGATCTGAATTC, oligo-2: GAATTCAGATC, annealed to make a 15-µM solution), 0.5 µl of T4 DNA ligase (New England Biolabs; catalogue no. 202L), and distilled water in a 50.2-µl total volume. The ligation reaction is allowed to continue overnight at 16°C. After the reaction, the DNA is purified by ethanol precipitation and dissolved in 25 µl of dH_2O.

PCR. The DNA that has been ligated to the linker oligo is mixed with 4 µl of 10× ThermoPol reaction buffer (New England Biolabs; catalogue no. B9004S), 4.75 µl of ddH_2O, 5 µl of 10× dNTP mix (2.5mM each dATP, dTTP, dGTP, and dCTP), 1.25 µl of oligo-1 (40 µM stock) in a final volume of 40 µl in a 500-µl-thin-wall Eppendorf PCR tube. The tube is first incubated at 55°C for 2 min on a thermal cycler; then 10 µl of an enzyme mix [8 µl of dH_2O, 1 µl of Taq DNA polymerase (5 units/µl), 1 µl of ThermalPol reaction buffer, and 0.025 unit of Pfu polymerase (Stratagene; catalogue no. 600250-51)] is added. Subsequently, the following PCR cycle is performed:

step 1: 72°C for 5 min;
step 2: 95°C for 2 min;
step 3: 95°C for 1 min;
step 4: 60°C for 1 min;
step 5: 72°C for 1 min;
step 6: go to step 3 for 22 times;
step 7: 72°C for 5 min;
step 8: 4°C indefinitely.

After the PCR, the DNA is purified using the Qiaquick PCR purification kit (Qiagen; catalogue no. 28106) and eluted in 60 µl of elution buffer provided with the kit.

Labeling of DNA and Microarray Hybridization

Labeling of Amplified DNA. In a colored Eppendorf tube, 200 ng of DNA from the previous step is mixed with 20 µl of 2.5× random primer solution (BioPrime kit,

Invitrogen; catalogue no. 18094-011) and dH_2O in a final volume of 42.5 μl. The mixture is boiled for 5 min and then cooled on ice for 5 min. Subsequently, 5 μl of 10× low dCTP mixture (2.5 mM each for dATP, dTTP, and dGTP, and 0.6 mM for dCTP), 1.5 μl of Cy5-dCTP (Amersham; catalogue no. PA55021) or Cy3-dCTP (Amersham; catalogue no. PA53021), and 40 units of Klenow DNA polymerase are added to tube. The tube is incubated at 37°C for 2 h. After the reaction, the labeled DNA is purified using the Qiagen PCR kit (Qiagen; catalogue no. 28106).

DNA Microarray Hybridization. In a new Eppendorf tube, 2.5 μg of Cy5-labeled chromatin immunoprecipitation (ChIP) DNA is mixed with 2.5 μg of Cy3-labeled genomic DNA and 36 μg of human Cot-1 DNA (Invitrogen; catalogue no. 15279-011). Sodium acetate(1/10 vol of 3 M) is added to the tube along with 2 vol of ethanol. After ethanol precipitation, the DNA is dissolved in 22.4 μl of hybridization buffer 1 (2.2× SSC/0.22% SDS). The mixture can be heated for 10 min at 37°C to facilitate the DNA to dissolve. Then 20 μl of hybridization buffer 2 (70% formamide/3× SSC/14.3% dextran sulfate) is added to the mixture, and the tube is first heated at 95°C for 5 min to denature DNA, then incubated at 42°C for 2 min. At this step, 4 μl of yeast tRNA (Sigma; catalogue no. R9001 at 10 μg/μl) and 3 μl of 2% BSA are added to the mixture, which is then spotted to a DNA microarray slide that has been incubated with the prehybridization solution for 40 min at 42°C. A 25-mm × 60-mm cover slip is then gently placed on top of the sample, and the hybridization is carried on in a hybridization chamber (Corning; catalogue no. 07-200-271) at 60°C overnight in a water bath.

Washing Microarrays. After the hybridization, the microarray slide is washed once with washing buffer 1 (2× SSC/0.1% SDS) at 60°C for 5 min in a glass slide staining dish. This is followed by a wash with buffer 2 (0.2× SSC/0.1% SDS) for 10 min at room temperature and three times with buffer 3 (0.2× SSC) at room temperature. The slide is then dried by a brief spinning at 1,000 × *g* in a table-top centrifuge.

Microarray Analysis and Identification of *In Vivo* DNA-Binding Sites

Microarray Scanning and Initial Analysis. The microarray slides are scanned using GenePix 4000B scanner from Axon Instruments (Foster City, CA), and each microarray image is first analyzed with the image analysis software GENEPIX PRO 3.0 to derive the Cy3 and Cy5 fluorescent intensity and background noise for all the spots on the array. The intensities for both Cy3 and Cy5 channels are first adjusted by subtracting the background intensity of each spot by using the formula $I_{channel} - B_{channel}$, where channel represents either Cy5 or Cy3. The intensities are further adjusted by subtracting the median intensity of all the blank spots on the array. If any of these values was lower than 10, it was then raised to 10.

Normalization. The normalization factor is calculated based on the intensities of spots that are considered good (more than 65% of pixels have intensities higher than the background intensity plus 1 SD). The median of the intensity ratios, I_{635}/I_{532} is then used to adjust the Cy3 channel intensity to the same level as that of the Cy5 channel.

Error Model. The quantitative amplification of small amounts of DNA generates some uncertainty in values for the low intensity spots. In order to track that uncertainty and average repeated experiments with appropriate related weights, a single array error model (1) is used. The significance of a measured ratio at a spot is defined by a statistic *X*, which is formulated as:

$$X = \frac{a_2 - a_1}{\sqrt[2]{\sigma_1^2 + \sigma_2^2 + f^2(a_1^2 + a_2^2)}} \qquad [1]$$

where $a_{1,2}$ are the intensities measured in the two channels for each spot, $\sigma_{1,2}$ are the uncertainties due to background subtraction, and *f* is a fractional multiplicative error such as would come from hybridization nonuniformities, fluctuations in the dye incorporation efficiency, scanner gain fluctuations, etc. The distribution of *X* can be found to be close to Gaussian distribution in experiments where Cy3 and Cy5 samples are identical. The significance of a change of magnitude *X* is then calculated using a one-sided probability model as follows:

$$P = 1 - Erf\left(\frac{X - \mu}{\sigma}\right) \qquad [2]$$

where μ is the average of *X*, and σ is the standard deviation of *X*. Because the intensities are normalized, μ should be near 0. The *Erf(x)* function is the standard normal accumulative distribution function corresponding to standard normal curve areas.

If the Cy3 and Cy5 samples are not identical, the Gaussian distribution can be skewed, because the ChIP can result in many DNA spots with significantly higher intensities in one channel than in the other. Since the input DNA is always present, the intensity distribution of those nonenriched spots can be used to obtain the parameters of *X* distribution. First, the spots whose *X* is a value less than 0 are identified. These spots should be on the left half of the Gaussian distribution. Their mirror spots are then generated by using $X^+ = -X^-$. The mean and SD of this new *X* can be calculated.

Combining Replicate Datasets. When the genomic binding sites of a protein are investigated using the above method, it is routine to perform several independent experiments so that the results are more reliable. To combine replicate data sets, each sample is first analyzed individually using the above single array error model. The average binding ratio and associated *P* value from these multiple experiments are then calculated using a weighted averaging analysis method (1). For each spot, the uncertainty in the *log*(*Ratio*) is defined as

$$\sigma = \log(a_2 / a_1) / X_{norm} \qquad [3]$$

Where $a_{1,2}$ are the intensities measured in the two channels for each spot, and X_{norm} is the normalized *X* for each spot. The weights of each spot are then defined as

$$w_i = 1/\sigma_i^2 \qquad [4]$$

The averaged *log*(*Ratio*) is then calculated using the following formula

$$\bar{x} = \frac{\sum_{i=1,n} w_i x_i}{\sum_{i=1,n} w_i} \qquad [5]$$

Where n is the total number of experiments, x_i and w_i are the *log*(*Ratio*) and weight for each experiment of a particular spot. $\bar{x}$ is normally distributed, and the averaged P value is then calculated using Eq. **2**, where the variable for $Erf(x)$ function is $\bar{x}$ instead. Target genes are selected based on a significant P value (e.g., <0.001).

1. Roberts, C. J., Nelson, B., Marton, M. J., Stoughton, R., Meyer, M. R., Bennett, H. A., He, Y. D., Dai, H., Walker W. L., Hughes, T. R., *et al*. (2000) *Science* **287,** 873–880.